

INSPECTING INSTRUMENT FOR PRINTED MATTER

Publication number: JP6044404

Publication date: 1994-02-18

Inventor: HIRASAWA TOSHITAKE

Applicant: TOKYO SHIBAURA ELECTRIC CO

Classification:

- international: G06K9/03; G06K9/20; G06T1/00; G06K9/20;
G06K9/03; G06K9/20; G06T1/00; G06K9/20; (IPC1-7):
G06K9/20; G06F15/62; G06F15/70; G06K9/03

- european:

Application number: JP19920197209 19920723

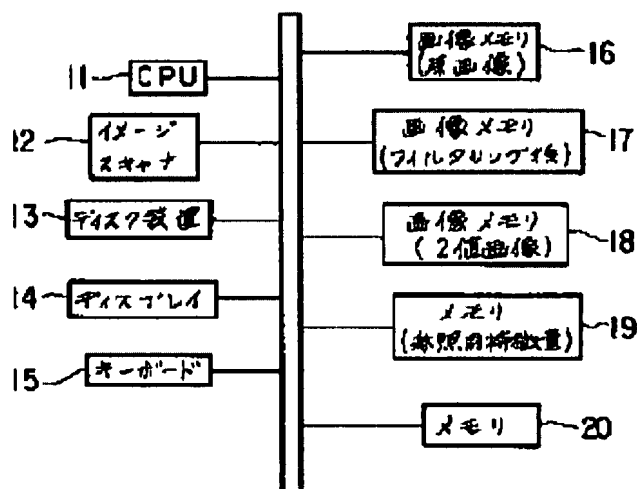
Priority number(s): JP19920197209 19920723

Report a data error here

Abstract of JP6044404

PURPOSE: To provide the inspecting instrument for printer matter, which sets a binarized threshold for extracting a character-string to an optimal value, even in the case a density difference between a character part and a background part is small, extracts only the character-string with high precision, and can execute an inspection with high precision.

CONSTITUTION: Image data on printer matter is read by a scanner 12, and stored in an image memory 16. By gradation image data in the image memory 16, first of all, a reference character-string in the left upper part of the printed matter is extracted, its feature amount is calculated, and with regard to the extracted reference character-string, the character recognition is executed. Subsequently, with regard to the character-string in the right lower part in which it is difficult to extract a character, the extraction is executed by using a feature amount of the reference character-string stored in a memory 19, and the character recognition is executed. In such a way, the extraction and the recognition processing of the character-string in which reflection lightness of a character and a pattern are separated, and the character-string in which reflection lightness is near are executed, and thereafter, those results of recognition are collated with each other, and only when both the same character-strings coincide with each other, a decision of validity for making printed matter effective is executed.



Data supplied from the esp@cenet database - Worldwide

(19)日本国特許庁(JP)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平6-44404

(43)公開日 平成6年(1994)2月18日

(51)Int.Cl. ⁵	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 K 9/20	3 4 0 J			
G 0 6 F 15/62	4 1 0 A	9287-5L		
15/70	3 2 0	9071-5L		
G 0 6 K 9/03		C 8623-5L		
		D 8623-5L		

審査請求 未請求 請求項の数1(全 7 頁)

(21)出願番号 特願平4-197209

(22)出願日 平成4年(1992)7月23日

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 平沢 利勇

神奈川県川崎市幸区柳町70番地 株式会社

東芝柳町工場内

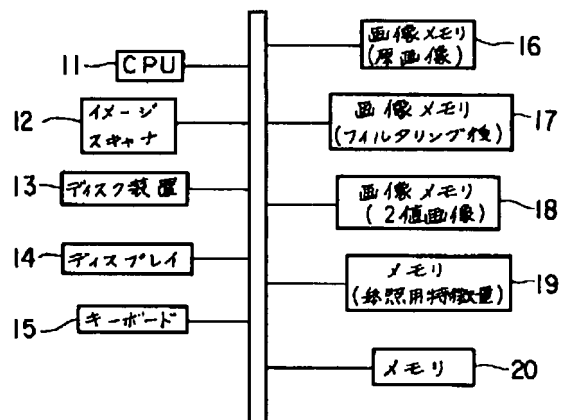
(74)代理人 弁理士 鈴江 武彦

(54)【発明の名称】 印刷物の検査装置

(57)【要約】

【目的】文字部と背景部との濃度差が小さい場合においても、文字列抽出のための2値化のしきい値を最適値に設定でき、文字列のみを高い精度で抽出して高精度の検査が可能となる印刷物の検査装置を提供する。

【構成】スキャナ12で印刷物上の画像データを読み取り、画像メモリ16に格納する。画像メモリ16内の濃淡画像データにより、まず、印刷物の左上部の参照用文字列を抽出して、その特徴量を算出し、抽出した参照用文字列について文字認識を行なう。次に、文字抽出が困難な右下部の文字列について、メモリ19に格納した参照用文字列の特徴量を用いて抽出を行ない、文字認識を行なう。こうして、文字と模様との反射明度が離れた文字列と、反射明度が近い文字列の抽出および認識処理を行なった後、それらの認識結果を互いに照合して、両同一文字列が一致したときにのみ印刷物を有効であるとする有効性の判定を行なう。



【特許請求の範囲】

【請求項 1】 2つ以上の同じ文字列画像を有する印刷物から画像データを収集する画像入力手段と、この画像入力手段によって収集された画像データから、前記 2つ以上の文字列のうち少なくとも 1つの文字列を抽出する第 1 文字列抽出手段と、この第 1 文字列抽出手段によって抽出された文字列の濃度的特徴量を算出する濃度的特徴量算出手段と、この濃度的特徴量算出手段によって算出された濃度的特徴量に基づき、その他の文字列を抽出する第 2 文字列抽出手段と、前記第 1 文字列抽出手段および第 2 文字列抽出手段によって抽出された前記 2つ以上の文字列をそれぞれ認識する認識手段と、この認識手段で認識された前記 2つ以上の文字列により前記印刷物の有効性を判定する判定手段とを具備したことを特徴とする印刷物の検査装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、たとえば、2つの同じ文字列画像を有する印刷物の有効性を検査する印刷物の検査装置に関する。

【0002】

【従来技術】従来、模様の上に文字が印刷されている印刷物における文字の印刷状態を検査する装置に関しては、たとえば、特開平 2-56688 号公報に示されるように、印刷文字の濃淡画像データから濃度ヒストグラムを作成し、その形状から適当なしきい値を決定し、2 値化処理を施した後、文字抽出を行っていた。

【0003】

【発明が解決しようとする課題】しかし、上記したような従来の方法では、文字と模様との反射明度が近い場合、濃度ヒストグラムにおけるそれぞれの濃度分布が分離しておらず、文字を十分に抽出し得る 2 値化のしきい値の決定が困難であり、そのため、高精度の検査が行えないという問題があった。

【0004】そこで、本発明は、文字部と背景部との濃度差が小さい場合においても、文字列抽出のための 2 値化のしきい値を最適値に設定でき、文字列のみを高い精度で抽出して高精度の検査が可能となる印刷物の検査装置を提供することを目的とする。

【0005】

【課題を解決するための手段】本発明の印刷物の検査装置は、2つ以上の同じ文字列画像を有する印刷物から画像データを収集する画像入力手段と、この画像入力手段によって収集された画像データから、前記 2つ以上の文字列のうち少なくとも 1つの文字列を抽出するとともに、その抽出した文字列の濃度的特徴量を求め、この求めた濃度的特徴量に基づき、その他の文字列を抽出する画像データ処理手段と、この画像データ処理手段によって抽

出された前記 2つ以上の文字列をそれぞれ認識する認識手段と、この認識手段で認識された前記 2つ以上の文字列により前記印刷物の有効性を判定する判定手段とを具備している。

【0006】

【作用】複数の同一文字列が複数箇所に印刷されている印刷物において、文字部と背景部との濃度差が小さい箇所に印刷されている文字列の抽出を行なう際、他の文字部と背景部との濃度差が大きく抽出容易な箇所に印刷されている文字列の抽出結果を用いて行なうことにより、文字部と背景部との濃度差が小さい場合においても、文字列抽出のための 2 値化のしきい値を最適値に設定して文字列のみを高い精度で抽出できる。したがって、印刷物の検査を高精度に行なうことが可能となる。

【0007】

【実施例】以下、本発明の一実施例について図面を参照して説明する。

【0008】図 1 は、本実施例に係る印刷物の検査装置の概略構成を示すものである。図において、11 は本装置全体の制御および各種画像処理を行なう CPU（センサ・プロセッシング・ユニット）であり、この CPU 11 には、印刷物から画像データを収集するイメージスキャナ 12、ディスク装置 13、ディスプレイ装置 14、動作の指示などを入力するキーボード 15、画像メモリ 16、17、18、参照用文字列における各文字の特徴量を格納するメモリ 19、および、初期値や基準値などを記憶するためのメモリ 20 が接続されている。

【0009】画像メモリ 16 は、イメージスキャナ 12 により収集される原画像データを格納するものであり、画像メモリ 17 は、画像入力系の高周波ノイズの抑制処理やエッジのコントラストを強調するためのフィルタリング処理を施したデータを格納するものである。画像メモリ 18 は、文字抽出に用いる 2 値画像データを格納するものである。メモリ 19 は、2 値化の際に用いる参照用文字列に関する特徴量（詳細については後述）を格納するものである。

【0010】図 2 は、検査対象となる印刷物の一例を示している。この印刷物 31 には、たとえば、同一の文字列が 2 箇所に印刷され、左上部に印刷された一方の文字列 32a は、文字と模様との反射明度の差が大きく（文字部と背景部との濃度差が大きく）、一般的に文字抽出し易い文字列であり、右下部に印刷された他方の文字列 32b は、文字と模様との反射明度の差が小さく（文字部と背景部との濃度差が小さく）、一般的に文字抽出し難い文字列であると仮定する。

【0011】次に、図 2 の印刷物 31 を例にとり、図 3 の印刷物検査の処理手順を示すフローチャートを参照して本発明の検査方法を説明する。まず、イメージスキャナ 12 によって印刷物 31 の画像データが読取られ、収集された濃淡画像データ（以下、濃度が低いことは明

度が低いこととする)は画像メモリ16に格納される(S1)。

【0012】画像データの収集が終了すると、まず、印刷物31の左上部の文字列32aを抽出して、その特徴量を算出する(S2)。この算出された特徴量はメモリ19に格納され、一方の文字列の文字抽出の参照データとなる。ここで、ステップS2における参照用文字列の抽出および特徴量算出の方法について説明する。

【0013】図4は、上記参照用文字列の抽出および特徴量算出を行なう際の処理手順の概略を示すフローチャートである。まず、抽出しようとする文字列を含む大まかな領域(たとえば、 $i \times j$ 画素)を、キーボード15を用いて指定する(S8)。そして、画像メモリ16内の濃淡画像データに対し、上記指定領域内の画像データについて濃度ヒストグラムを作成すると、一般に図5に示すように、文字部および模様部を示す2つの山が現われる。ここで、たとえば濃度ヒストグラムを解析することにより、2つの山(極大値)間でのヒストグラムの極小値を与える濃度値 t_m を求め(S9)、この t_m によって濃淡画像データをしきい値処理することにより、左上部の文字列32aを2値画像として抽出する(S10)。

【0014】次に、抽出した2値化データについて、横方向および縦方向にそれぞれ抽出画素で累積することにより、図6に示すような周辺分布を求める(S11)。横方向に累積した周辺分布から、たとえば、「非0」分布の範囲(つまり、文字画像に係る画素が1画素でも存在する範囲)を文字領域の縦方向とし、また、同様に縦方向に累積したときの分布から、たとえば、複数の「非0」分布の範囲を各文字の横方向として、図7に示すように各文字の領域を検出して文字抽出を行なう(S12)。

【0015】文字の抽出が終了すると、後の2値化処理で用いる文字の特徴量として、各文字における文字部の画素数、および、それらの総和である文字列の全画素数を計数し、それを文字の特徴量としてメモリ19に格納する(S13)。

【0016】このようにして、参照用文字列32aの抽出、および、その文字の特徴量(画素数)の算出を行なうと、次に、抽出された参照用文字列32aの各文字について、辞書パターンとの比較処理を行なうことにより文字の認識を行ない、その認識結果をメモリ20に格納する(S3)。

【0017】参照用文字列32aの認識が終了すると、次に、文字と模様との反射明度が近いために、文字抽出が困難な右下部の文字列32bについて、メモリ19に格納した左上部の参照用文字列32aの特徴量を用いて文字抽出を行なう(S4)。ここで、ステップS4における文字列の抽出方法について説明する。

【0018】図8は、上記文字列の抽出を行なう際の処

理手順の概略を示すフローチャートである。まず、画像メモリ16に格納された濃淡画像データから、ステップS8と同様に、抽出しようとする文字列を含む大まかな領域(たとえば、 $i \times j$ 画素)を、キーボード15を用いて指定する(S14)。

【0019】次に、上記指定された領域内の画像データに対してフィルタリング処理を行ない、その結果を画像メモリ17に格納する(S15)。このフィルタリング処理の一例は、たとえば、図12に示すような 3×3 マスクの空間フィルタを用いた平滑化であり、図9に示すように、文字列の背景部に存在する高周波成分を含む斜線などの繰返しパターンPを除去するのに有効である。

【0020】フィルタリング処理が終了すると、画像メモリ16内の濃淡画像データのうち、上記指定領域内の画像データについて濃度ヒストグラムを作成し、文字領域を切出すためのしきい値を決定する。ところが、図11に示すように、右下部の文字列32bは、文字と模様との反射明度が近いため、濃度ヒストグラムが1つの山になってしまう。そこで、文字列32aと文字列32bの文字部全画素数がほぼ一致し、背景部に比べて文字部の濃度が低いという前提で、先にメモリ19に格納しておいた文字列32aの文字部全画素数を用いて、ステップS4で指定した全画素数($i \times j$)に対する文字列32aの文字部全画素数の占める割合 $r\%$ を算出し、濃度ヒストグラムの累積値の低い方から $r\%$ になる濃度値を求める(S16)。

【0021】こうして求めた濃度値をしきい値として文字列32bの2値化を行ない、その結果を画像メモリ18に格納しておく(S17)。この後、文字列32aの2値画像から周辺分布および各文字領域を求めたステップS11、S12と同様の手順で、画像メモリ18に格納した2値画像から図10に示すような周辺分布を求め(S18)、各文字領域を検出する(S19)。

【0022】このとき、前記一様なしきい値により得られる2値画像は、たとえば、図13(a)に示すように、文字「M」のようなかすれた文字を含んでしまう。そこで、文字かすれをなくすべく、しきい値レベルを上げると、図13(b)に示すように、文字「682」の背景にノイズ成分Nが重なってしまう。すなわち、ステップS17のような全領域を一様なしきい値で2値化することでは、十分な文字抽出ができない。したがって、ステップS15で格納したフィルタリング後の濃淡画像データに対して、さらに細かい領域、たとえば、1文字ごとにしきい値を再設定して文字抽出を行なう。

【0023】すなわち、ステップS16と同様に、メモリ19に格納した文字列32aの各文字の画素数を用いて、各文字領域の全画素数に対する各文字部画素数の占める割合 $r_1, r_2, \dots, r_n\%$ (n =文字数)を計算し、各文字領域における濃度ヒストグラムの累積値が濃度の低い方から $r_1, r_2, \dots, r_n\%$ になる各濃度値を求める(S

20)。これらの濃度値を各文字ごとのしきい値として各文字領域ごとに2値化処理を行ない、その結果を画像メモリ18に格納して文字抽出を終了する(S21)。

【0024】このようにして、文字列32bの抽出を行なうと、次に、文字列32aと同様に、文字列32bの各文字について文字認識を行ない、その結果をメモリ20に格納する(S5)。

【0025】以上のようにして、印刷物31に印刷された文字と模様との反射明度が離れた文字列32aと、文字と模様との反射明度が近い文字列32bの抽出、および、認識処理を行なった後、最後に2つの同一文字列32a、32bに対する認識結果を互いに照合して(S6)、たとえば、2つの同一文字列32a、32bの各文字認識結果が全て一致したときのみ、印刷物31を有効であるとするなどの、印刷物31の有効性の判定を行なう(S7)。

【0026】このように、たとえば、同一文字列が2箇所に印刷されていて、一方の文字列は文字と模様との反射明度の差が大きく、他方の文字列は文字と模様との反射明度の差が小さい印刷物において、文字と模様との反射明度の差が小さい文字列を2値化して抽出する際、文字と模様との反射明度の差が大きい文字列の2値化結果を用い、各文字ごとの文字部画素数および文字部全画素数を参照して2値化を行なうことにより、文字と模様との反射明度の差が小さい場合においても、文字列抽出のための2値化のしきい値を最適値に設定して文字列のみを高い精度で抽出できる。したがって、印刷物の有効性検査などを高精度に行なうことが可能となる。

【0027】なお、前記実施例では、文字と模様との反射明度の差が大きい文字列、および、文字と模様との反射明度の差が小さい文字列の、それぞれ1箇所ずつ同一文字列が印刷された場合について説明したが、上記場合に加えて、文字と模様との反射明度の差が小さい文字列が複数個ある場合など、3つ以上の同一文字列が印刷されていても、少なくとも1箇所に文字と模様との反射明度の差が大きい文字列が印刷されていれば、本発明が適応できる。

【0028】また、前記実施例では、全画素数($i \times j$)に対する文字列32aの文字部全画素数の占める割合 $r\%$ 、および、各文字領域の全画素数に対する各文字部画素数の占める割合 $r_1, r_2, \dots, r_n\%$ (n =文字数)を用いているが、これらの割合 $r\%$ 、および、 $r_1, r_2, \dots, r_n\%$ は、入力画像データおよび参照用文字列の2値

化レベルなどの条件によって変動させることもできる。

【0029】また、前記実施例では、ある参照する文字列を構成する、それぞれの文字の特徴情報として、文字部の画素数(面積)、および、文字部の濃度が模様部よりも濃いことを用いているが、これに限らず、たとえば、文字の周囲長、文字幅、重心、連結成分数などの幾何学的特徴量、および、文字部の濃淡画像データの分散などの濃度的特徴量を採用、および、組合わせて、2値化しきい値の決定に用いることができる。

【0030】さらに、前記実施例では、濃淡画像データを用いた場合について説明したが、カラー画像データの入力を行ない、色を構成する、たとえばR、G、Bの画像データを用いた場合においても、各種変換を行なった結果を濃淡画像データとして扱い、前記列挙した特徴情報を2値化しきい値の決定に用いることができる。

【0031】

【発明の効果】以上詳述したように本発明によれば、文字部と背景部との濃度差が小さい場合においても、文字列抽出のための2値化のしきい値を最適値に設定でき、文字列のみを高い精度で抽出して高精度の検査が可能となる印刷物の検査装置を提供できる。

【図面の簡単な説明】

【図1】本発明の一実施例に係る印刷物の検査装置の構成を概略的に示すブロック図。

【図2】検査対象印刷物の一例を示す平面図。

【図3】印刷物検査の処理手順を示すフローチャート。

【図4】参照用文字列の抽出および特徴量算出の処理手順を示すフローチャート。

【図5】参照用文字列の濃度ヒストグラムを示す図。

【図6】周辺分布の算出を説明する図。

【図7】文字領域の抽出を説明する図。

【図8】文字列抽出の処理手順を示すフローチャート。

【図9】フィルタリング処理の効果を説明する図。

【図10】周辺分布の算出を説明する図。

【図11】文字部の濃度ヒストグラムとしきい値との関係を示す図。

【図12】フィルタリング処理に用いる 3×3 マスクの一例を示す図。

【図13】文字の2値化例を示す図。

【符号の説明】

11…CPU、12…イメージスキャナ、16、17、18…画像メモリ、19、20…メモリ、31…印刷物、32a、32b…文字列。

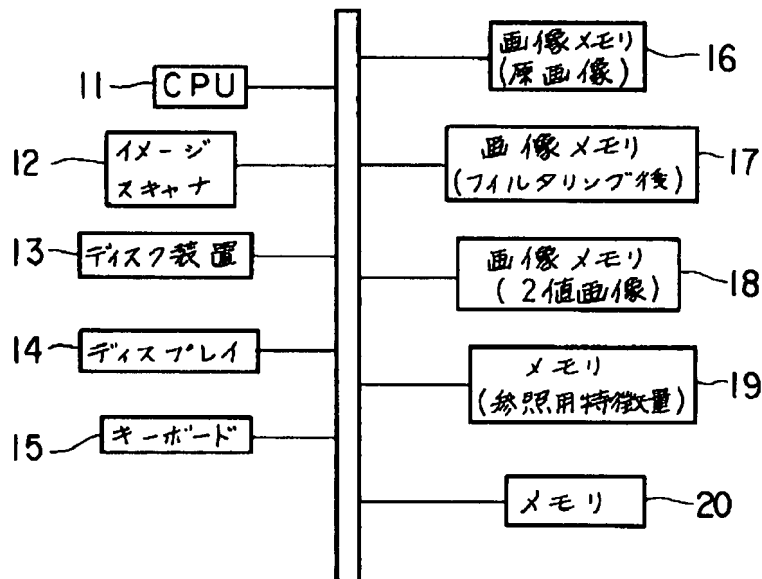
【図7】

M 6 8 2

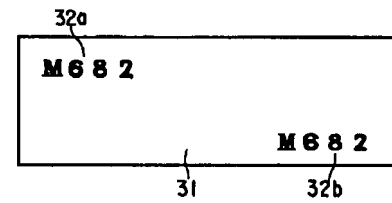
【図9】



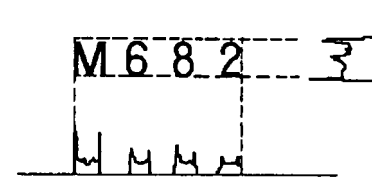
【図1】



【図2】

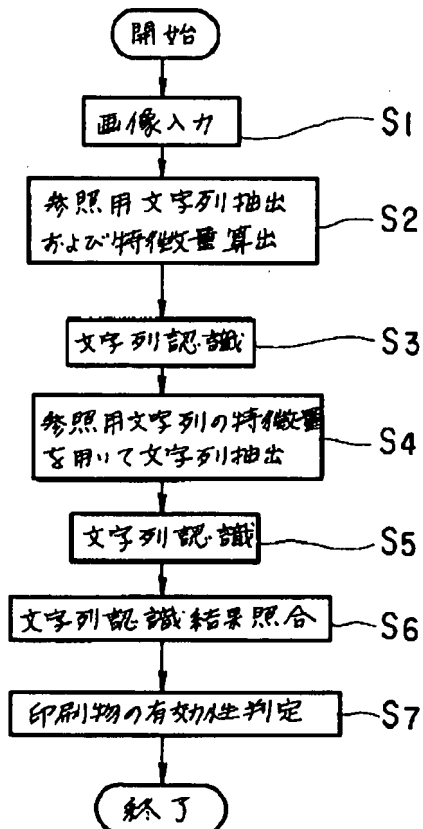


【図6】

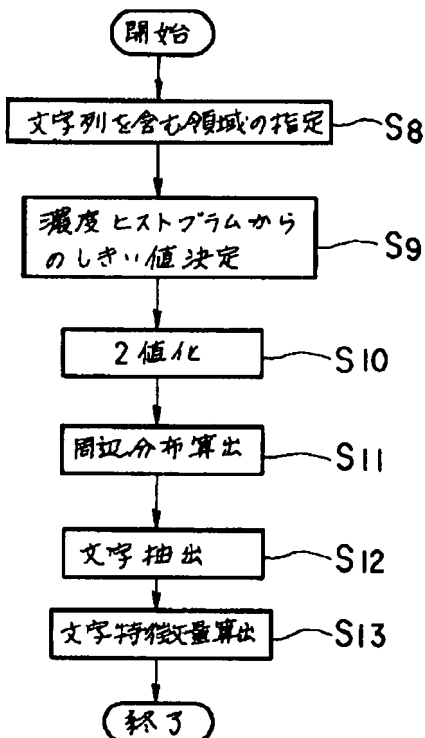


【図12】

【図3】



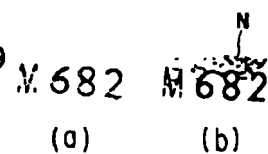
【図4】



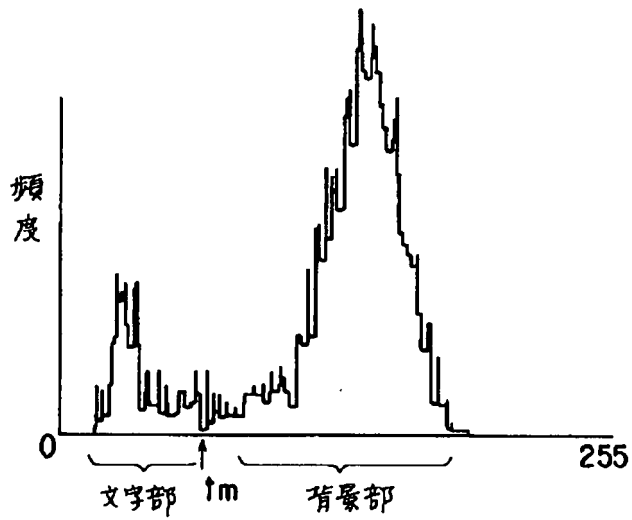
1	1	1
1	1	1
1	1	1

 $\times \frac{1}{9}$

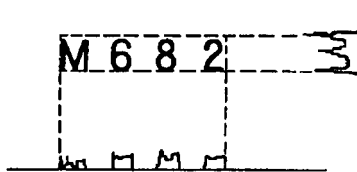
【図13】



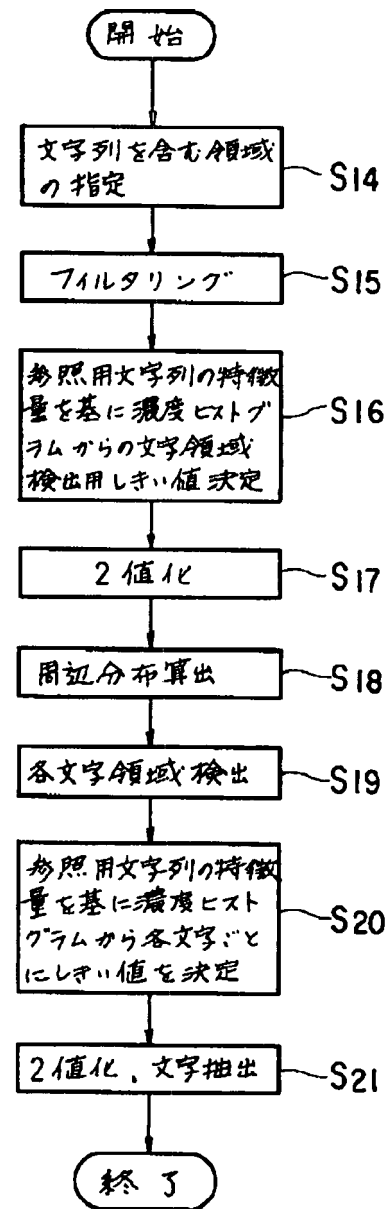
【図5】



【図10】



【図8】



【図 1 1】

